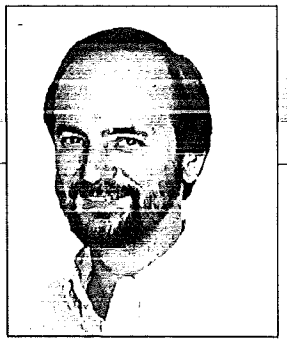
3481 p22

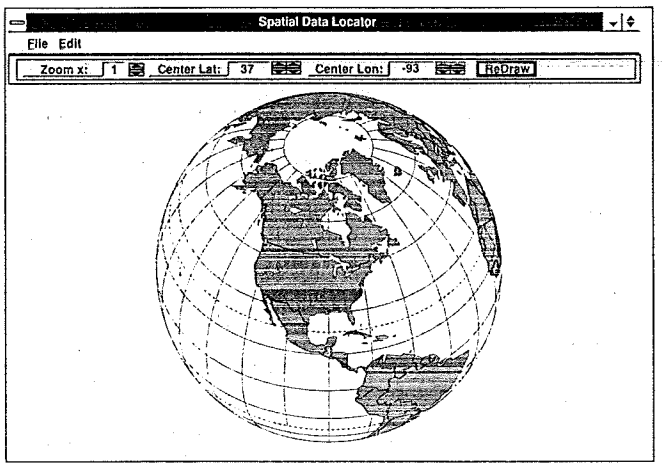


By Eric J. Strand

Wide Area Information Servers Provide a Wealth of Information

Eric J. Strand is GIS project leader at Management Assistance Corporation of America, Inc., an information system engineering company. He can be reached at Management Assistance Corporation of America, Inc., 2629 Redwing Road, Fort Collins, CO 80526, USA.

The global information cornucopia is overflowing with billions and billions of data bytes, providing a wealth of information and compounding its growth every minute. Many computer companies, government agencies and universities have created a consortium to define an open standard for data access. To simplify access to computerized information, the consortium is developing public-domain software based on the Information Retrieval Service Definition and Protocol Specification (Z39.50). This standard protocol gave birth to Wide Area Information Servers (WAIS), which are registered in a Directory of Servers maintained on the Internet by Thinking Machines Corp., Cambridge, Mass., USA. The servers are capable of responding to information retrieval requests conforming to the Z39.50 standard. All Z39.50 clients and servers support text search and retrieval, and many support other document types such as graphics, hypertext and video.



A U.S. Geological Survey mapping interface addition to WAIS enables spatial searches for data and information.

The information servers listed by the Directory of Servers offer access to more than 400 databases around the world. Many publicly available servers can be used free. For example, Thinking Machines maintains the Central Intelligence Agency's World Factbook, weather maps and forecasts, patents, WAIS software and documentation, and the Directory of Servers. COSMIC, NASA's software distribution center, maintains a catalog of government-sponsored, public-domain software packages. The National Institute of Health publishes research grant opportunities. The Library of Congress is creating a server for its card catalog. Also, the Internet Archie facility maintains directories of software and data available from networked host computers. Archie only helps you find *where* a file is. WAIS helps you find *what* the file contains.

The U.S. Geological Survey (USGS) maintains the Earth Science Data Directory (ESDD) and is extending WAIS to accommodate the directory. USGS will register a sub-directory of servers for ESDD and list information sources from other organizations. Current sources include the interagency Global Change Master Directory, the National Oceanic and Atmospheric Administration (NOAA) National Environmental Data Referral Service, the Department of Energy Climate Change Directory, the USGS Arctic Environment Data Directory (a subset of ESDD), the USGS Water Data Abstracts and the USGS Spatial Data Clearinghouse. This latter information source is a directory to other sources such as the Distributed Spatial Data Library, the Geographic Names Information System, the Map Chart Information System, the Cartographic Catalog and the Aerial Photo Summary Records Systems.

How Does WAIS Work?

The WAIS client/server implementation overcomes traditional barriers to information access by eliminating the need for data to be converted to different formats or to conform to a single presentation style. WAIS server software maintains information in suitable formats, and WAIS client software provides appropriate information retrieval and presentation capabilities.

WAIS converts search words to the Z39.50 standard information retrieval protocol and sends the converted search request to each server associated with the selected information source. Each server then matches the search request to the contents of its files and databases. The results are sent from each server and presented to the user as a consolidated list of all document titles that satisfy the request. The user then can select a title and have the document sent across the network from the respective information server.

Z39.50 lists documents based on relative scores determined by the information server that ranks the document according to its probable relevance to the user. Scoring documents based on nouns and other words enables the user to talk to WAIS in plain English. By avoiding the need to remember key words and providing "relevance feedback" through scoring, a WAIS user does not need sophisticated skills or specific technical training to find the information he or she wants.

USGS is adding several interesting capabilities to WAIS (Christian and Gauslin), including phrase searching, key word searching within fields and location searching. Location searching uses a graphically displayed map, which could be the entire globe (see accompanying figure), to help find and select a location. The extended capabilities for WAIS are designed within the constraints of the Z39.50 standard to allow users of the USGS/WAIS client software to access any WAIS server or use the extended capabilities for USGS servers. Thus, the user can easily navigate among the various systems.

The global change data management community may use WAIS with the interagency Global Change Master Directory, which is the source for references to key global change data and information. The International Council of Scientific Unions' Committee on Data also is considering WAIS for developing a Directory of Directories, which

would provide a general gateway to available government information.

Are There Other Ways?

Archie, Gopher and WorldWide Web are other popular client-server tools within the Internet community for locating and retrieving information across the net (LaQuey and Ryer, 1993). Like WAIS, Archie and Gopher provide the user with an overview of likely places to find the desired information and then help the user locate the specific information items. WorldWide Web allows the user to access hypertext links with a simple mouse click. As each hypertext link is activated, World-Wide Web helps the user blaze a trail across the "web" of Internet networks by opening network connections to referenced information as needed.

The successful cooperative development and use of WAIS demonstrates the potential value of the tool for corporate information retrieval. USGS has expanded this vision of information retrieval to support spatial queries for geographical information. Providing a new way to use ESDD is a pioneering venture that promises simple, effective access to a world of geographical data and information. As a result, new doors will open for transferring knowledge between global researchers, educators and the public.

The Clearinghouse for Networked Information Discovery and Retrieval (CNIDR), Research Triangle Park, N.C., USA, funded in part by the National Science Foundation, is working closely with the developers of client-server tools to evolve consistent and compatible tools for locating and retrieving information accessible through Internet. To demonstrate WAIS from CNIDR, log in as wais and execute quake.think.com.

For more information about WAIS and CNIDR, contact George Brett, ghb@concert.net, or Jane Smith, jds@concert.net, CNIDR, Center for Communications - MCNC, PO Box 12889, Research Triangle Park, NC 27709-2889, USA [919-248-1886, FAX: 919-248-1405]. For more information on WAIS map location extensions, contact Timothy Gauslin, tgauslin@is-dres.er.usgs.gov, U.S. Geological Survey, 802 National Center, Reston, VA 22092, USA [703-648-5980].

References

Christian, Eliot J. and Timothy L. Gauslin. **Wide Area Information Servers: Standards-Based Access to Information and Data**, *American Chemical Society Symposium Proceedings* (in preparation).

LaQuey, Tracy and Jeanne C. Ryer. 1993. *The Internet Companion: A Beginner's Guide to Global Networking*, Addison-Wesley Publishing Co., March.

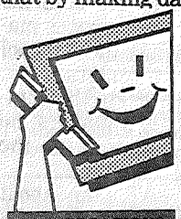
c., 155 E. Boardwalk Drive, Suite 250, Fort Collins, CO 80525, USA

3481

In search of easier access to databases

Commercial online databases are a lot like MS-DOS: they let you do all sorts of interesting things, but sometimes it seems like you need a master's degree in computer science to figure them out. Indeed, there is a group of people that makes a comfortable living searching through these information sources for people who have neither the time nor the money to decipher them.

But four companies are working on a project that could change all that by making databases nearly idiot-proof.



CONVERSATIONS WITH FRED

Thinking Machines of Cambridge, Apple Computer, Dow Jones & Co. (already heavily involved in the online database business) and KPMG Peat Marwick, are working on something called a "wide-area information server" or WAIS.

The goal is to develop a universal interface that doesn't require a manual the size of the Manhattan phone book, and which would let somebody automatically search through a whole series of databases with just one set of keystrokes, according to a paper on the subject by Brewster Kahle of Thinking Machines.

As an example of how this concept could change things, let's say you're writing *the* article on IBM. Today, you might call up Dow Jones to retrieve any recent financial news about the company, using one set of search instructions. Then you might call up Lexis and use another set of search instructions to see if they have been involved in any recent court battles. To round out your search, you might dial up another database, Nexis. And on and on and on.

With a WAIS system, you would just call up your own WAIS program and type "IBM." The program would call up all these databases for you (or at least the ones you specified beforehand, in a set-up file), translate your request into the code each individual program needs and get the information you need. No fuss, no muss.

There are three main components of a WAIS system. The first is the user interface, through which somebody can pose questions essentially in English, rather than in cryptic gobbledygook. The interfacing program then translates this into computerese, and using a set of universal communications and query protocols (part two), poses the question to the system's third component, the servers.

The servers are the computers which actually hold the data. They could be a computer on a user's company network or a mainframe halfway around the world reached by modem. When a server machine gets the request, it sorts through its data and sends back a list of documents its found that match that request to the interface. The user can then chose one of the documents to view, or can further narrow his search.

Automate such a system and you could let users create their own "personal newspapers" by having their computers dial into news services periodically and downloading any stories in categories in which they're interested.

The four companies are already experimenting with a national WAIS system through the Internet research network, whose users can now seek answers to questions in a variety of academic fields, based on information in public messages posted on the network in specific conferences.

As with the object-oriented programming Fred told you about last week, none of the basic ideas here are really new. For several years, CompuServe has offered a service that lets a user search a large number of databases through one interface. Its IQuest even has a "smart scan" feature that helps users narrow their search so they don't pay to access databases that may not have what they're looking for. CompuServe also has another service that lets users set up their own "news clipping" folder that automatically collects AP stories of interest to them.

What is different is that the four companies have made their protocols and database "hooks" public. As happened with DOS, this could lead to an explosion of WAIS-related products and programs — possibly at lower costs than those charged by most present databases. The protocols the four companies have developed allow for expansion into such areas as video and graphics so that one day, an art student writing a paper on Rembrandt, say, could call up a series of reproductions of his works, rather than having to trek to a library.

Fred sees some other potential refinements. Why not set up high and low-speed database links? People who need information *right this second* could pay a premium for that kind of access through a high-speed network. People who can wait a day or two for the data, though, would pay less to get the access through a commercial analog of Fidonet, an efficient but somewhat slow system for transmitting information. That way, you could open up online services to a far larger number of people and still make money — helping ease Fred's worry about the growing gap between the haves and the have-nots in the Information Age.

Fred the Middlesex News Computer eagerly awaits your call. With a computer and modem, you can call him, any time, day or night, at (508) 872-8461. Set your parameters to 8-1-N and up to 2400 baud. You can also send news tips or suggestions to Fred via CompuServe to 73727,545 or via Internet to adamg@world.std.com.

LUNCH MENUS

● FRAMINGHAM

ELEMENTARY
Alternate choice sandwich lunch offered daily
TUESDAY: Spaghetti, meat sauce, vegetables, bread & butter, fruit.
WEDNESDAY: Turkey fricassee, whipped potatoes, vegetables, bread & butter, chocolate pudding.
THURSDAY: Middle & senior high schools, in-service, no lunches served. Juice, tuna salad sandwich, chips, carrot sticks, ice cream.
FRIDAY: Double cheese pizza, salad, fruit cup.

● KEEFE TECH

TUESDAY: Juice, hot dog, salad, baked beans, fruit; ravioli w/sauce, bread & butter.
WEDNESDAY: Juice, meatballs on roll, potatoes, brownie or fruit; cheese croissant.
THURSDAY: In-service day, 11 a.m. dismissal.
FRIDAY: Juice, pizza or sandwich, salad, cookie or fruit.

● NATICK

ELEMENTARY AND MIDDLE SCHOOLS
Elementary schools-Lunch price $1.25. Choice of hot dogs, PB&J sandwiches, served with soup or juice, chips, vegetable, dessert or fruit offered daily.
Middle schools- Lunch price $1.50. Choice of pizza, subs, hot dogs, PB&J sandwiches, served with soup or juice, chips, vegetable, dessert or fruit offered daily.
TUESDAY: American chop suey, salad, French bread, fruit.

No KIDDING?

Two Civil War nurses

1. Louisa May Alcott
2. Walt Whitman

Source: World Features Syndicate

World Features Syndicate will pay $5 for your idea if published. Send c/o this newspaper to P.O. Box 660, Maple Shade, NJ 08052. Include name, address, telephone number and source.

WEDNESDAY: Turkey fricassee, mashed potato, gravy, peas, roll, sliced peaches.
THURSDAY: Hot meatball sub, carrots, chips, fruit.
FRIDAY: Chicken w/rice soup, double cheese pizza, salad, jello, sliced fruit.

● WSES

TUESDAY: Cream of broccoli soup, meatball sub, green beans, chocolate pudding.
WEDNESDAY: Brunswick stew, vegetables, boiled potato, roll, cookie.
THURSDAY: Oven breaded chicken, chantilly potato, vegetables, oatmeal bread, fruit.
FRIDAY: Breaded fish & cheese, mashed potato, peas, wheat bread, fruit.

BIRTHS

● FRAMINGHAM UNION HOSPITAL

JULY 7, 1991
FARLEY: A daughter, Kayla Ann Farley, to Lisa A. (Comerato) and Jim Farley of Hopkinton.

AUGUST 11, 1991
HOUGH: A daughter, Samantha Marie, to Angela Samaroco Hough and William Hough of Uxbridge.

GOKEY: A daughter, Olivia Alessandra, to Michelle and John Gokey of Marlboro.

AUGUST 13, 1991
DUPONTE: A daughter, Mindy Lee, to Irene and Jeff DuPonte of Norfolk, Maryland.

AUGUST 14, 1991
BLANEY: A son, Brendan Daniel, to Mary (Arnold) and Daniel Blaney of Holliston.

AUGUST 15, 1991
JOHNSON: A son, Robert Warren to Cynthia D. Lindstrom-Johnson and Harvey R. Johnson of Medway.
WALLACE: A daughter, Paige Leeann, to Kim Crosby-Wallace and Brian Wallace of Natick.
CARTER: A daughter, Samantha Marie, to MaryJane and David Carter of Southboro.
ATTIAS: A son, Revven Eric, to Debra and Moshe Attias of Framingham.
AUGUST 16, 1991
CAMPBELL: A son, Michael James, to Paula (Pasciuti) and Ronald Campbell of Hudson.
RUTKOWSKI: A son, Charles Edward, to Kathleen Burns Ratkowski and Gregory Allen Rutkowski of Hopkinton.
DASILVA: A son, Tiago Moura, to Luisa Dias Dasilva and Manuel Alves Dasilva of Framingham.
AUGUST 17, 1991
HANNA: A daughter, Kaitlyn Marie, to Bonnie Ann Elizabeth Hanna of Framingham.
MOORE: A son, John Joseph, to Deborah Baptista Moore and John D. Moore of Holliston.

18A-1090
CONNEXIONS
1.500
MONTHLY
NOVEMBER 1991

In addition, developers are faced with a number of architectural issues. The system must be scalable; that is, it must allow for the future growth of both the complexity and number of clients and servers. It must be secure; each server's data must be protected from corruption, and the privacy of the users must be ensured. Lastly, since an unreliable source is useless in a corporate environment, access must be thoroughly robust.

System overview

The prototype WAIS system takes advantage of current state-of-the-art technology, and presents solutions to all of the above problems. The system is composed of three separate parts: *Clients, Servers,* and the *Protocol* which connects them.

The Client is the user interface, the server does the indexing and retrieval of documents, and the protocol is used to transmit the queries and responses. The client and server are isolated from each other through the protocol. Any client which is capable of translating a users request into the standard protocol can be used in the system. Likewise, any server capable of answering a request encoded in the protocol can be used. In order to promote the development of both clients and servers, the protocol specification is public, as is its initial implementation.

On the client side, questions are formulated as English language questions. The client application then translates the query into the WAIS protocol, and transmits it over a network to a server. The server receives the transmission, translates the received packet into its own query language, and searches for documents satisfying the query. The list of relevant documents are then encoded in the protocol, and transmitted back to the client. The client decodes the response, and displays the results. The documents can then be retrieved from the server.

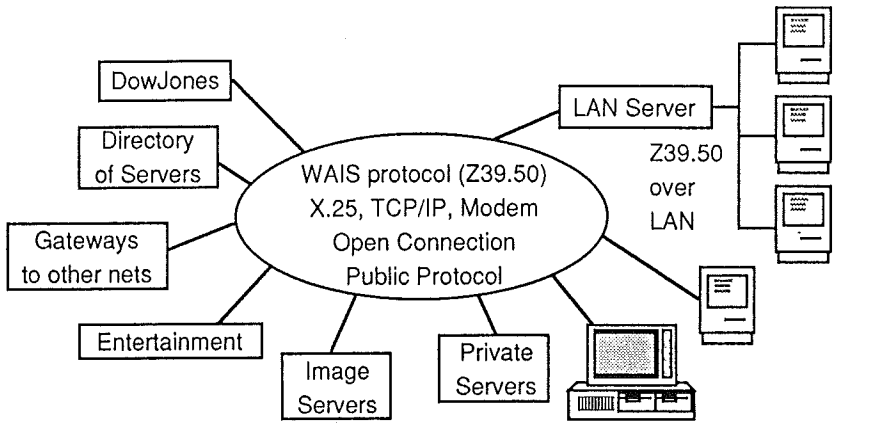


Figure 1: WAIS System Components

Digital researcher

The traditional information research scenario is familiar to anyone who has ever visited a reference desk at a public or corporate library. The client approaches a librarian with a description of needed information. The librarian might ask a few background questions, and then draws from appropriate sources to provide an initial selection of articles, reports, and references.

continued on next page

Wide Area Information Servers *(continued)*

The client then sorts through this selection to find the most pertinent documents. With feedback from these trials, the researcher can refine the materials and even continue to supply the user with a flow of information as it becomes available. Monitoring which articles were useful can help keep the researcher on-track.

The WAIS system is an attempt at automating this interaction: the user states a question in English, and a set of document descriptions come back from selected sources. The user can examine any of the items, be they text, picture, video, sound, or whatever. If the initial response is incomplete or somehow insufficient, the user can refine the question by stating it differently.

In addition, the user may also mark some of the retrieved documents as being "relevant" to the question at hand, and then re-run the search. The server recognizes the marked documents, and attempts to find others which are similar to them. In the present WAIS system, "similar" documents are simply ones which share a large number of common words; however, there is potentially no upper limit on the intelligence of a server in determining what similarity entails. This method of information retrieval is called "relevance feedback." The idea has been around for many years [1] and the first commercial system utilizing it, *DowQuest* [2], was voted Database of the Year by *ONLINE Magazine* in January 1989.

User interfaces

Users interact with the WAIS system through the *Question* interface. The interface may appear different on various implementations: for example, a character display terminal will have a different look than one which is capable of displaying bit-mapped graphics. The key, however, is that the user need only become familiar with one interface which provides access to all available information sources.

The WAIS system, in this first incarnation, was designed to be used by accountants and corporate executives who are relatively untrained in search techniques. Consequently, to aid those users who have neither the time nor desire to learn a special purpose query language, the system uses English language queries augmented with relevance feedback. While the system's servers currently do not extract semantic information from the English queries, they do their best to find and rank articles containing the requested words and phrases. Used in conjunction with relevance feedback, this method of searching has proven to be more than adequate for the types of searches and databases typically encountered.

Several user interfaces are in use or under development at Thinking Machines, Apple Computer, Dow Jones, and elsewhere. As shown on the facing page, a typical search scenario has the following steps:

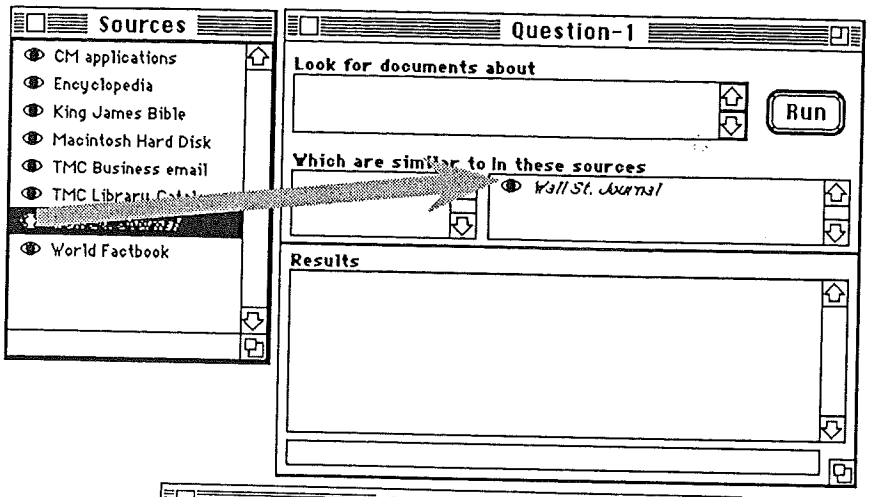
Step 1: Sources are dragged with the mouse into a Question Window. A question can contain multiple sources. When the question is run, it asks for information from each included source.

Step 2: When a query is run, headlines of documents satisfying the query are displayed.

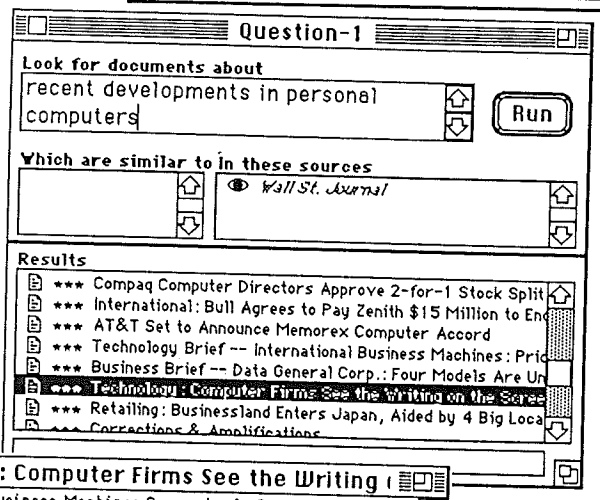
Step 3: With the mouse, the user clicks on any result document to retrieve it.

Step 4: To refine the search, any one or more of the result documents can moved to the "Which are similar to:" box. When the search is run again, the results will be updated to include documents which are "similar" to the ones selected.

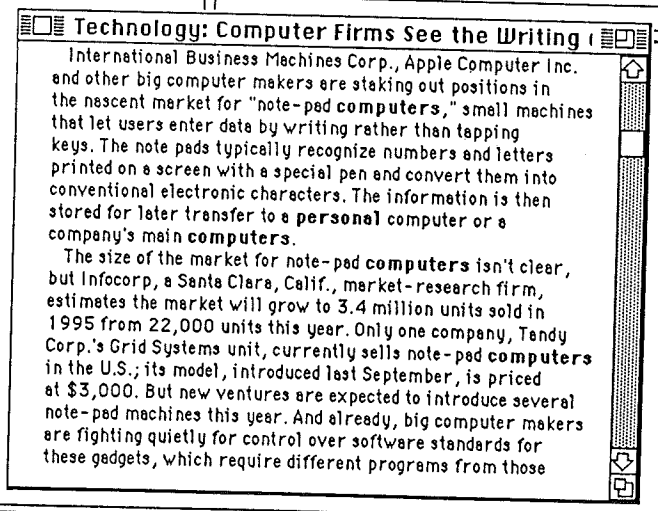
Step 1



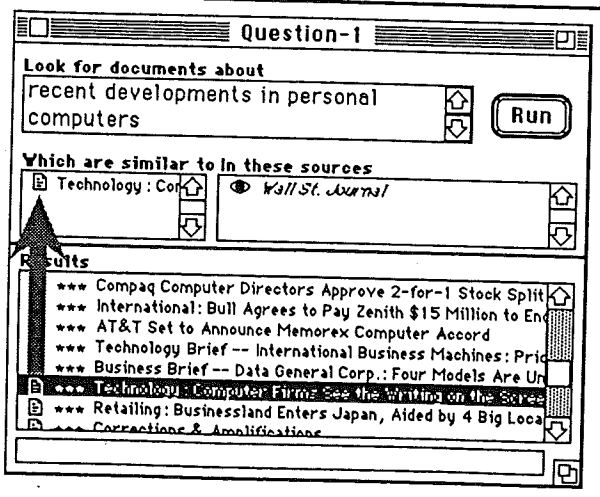
Step 2



Step 3



Step 4



continued on next page

Wide Area Information Servers *(continued)*

Contacting remote information sources

From the users point of view, a server is a source of information. It can be located anywhere that one's workstation has access to: on the local machine, on a network, or on the other side of a modem. The user's workstation keeps track of a variety of information about each server. The public information about a server includes how to contact it, a description of the contents, and the cost. In addition, individual users maintain certain private information about the servers they use. Users need to budget the money they are willing to spend on information from particular servers, they need to know how often and when each server is contacted, and they need to assess the relative usefulness of each server. This information helps guide the workstation in making cost effective decisions in contacting servers.

With most current retrieval systems, complications develop as soon as one begins dealing with more than one source of information. The most common problem is that of asking a particular question. For example, one contacts the first source, asks it for information on some topic, contacts the next source, asks it the same questions (most likely using a different query language, a different style of interface, a different system of billing), contacts the next source, and so on. One of the primary motivations behind the initial development of the WAIS system was to replace all this with a single interface.

With WAIS, the user selects a set of sources to query for information, and then formulates a question. When the question is run, the system automatically asks all the servers for the required information with no further interaction necessary by the user. The documents returned are sorted and consolidated in a single place, to be easily manipulated by the user. The user has transparent access to a multitude of local and remote databases.

A personal newspaper

In addition to providing interactive access to a vast quantity of information, the WAIS system can also be used as a rudimentary personal newspaper. A virtually unlimited number of queries can be saved, and updated at periodic intervals. To do this, the user's workstation is directed to contact each server at certain set times. When a source of information is contacted, any questions referencing that source are updated with new documents. The users can then easily browse through the results the next morning.

To make the ideal electronic personal newspaper, a system designer would need certain technologies which are not available today. Most computer screens are too small to allow efficient browsing of large amounts of text. Additionally, current data transmission speeds do not allow fast enough scanning if the text is not resident on the user's machine.

Despite current limitations, the WAIS system employs a number of features which will be found in the personal newspaper of the future:

- Clear displays of which questions have new documents
- Searches performed at night to hide communications delays
- Documents stored on disk for future reference
- Tools provided to quickly view stored documents

With these techniques, we have established a foundation of user support and acceptance.

Servers

The WAIS system was designed to be used by those who wish to sell information, as well as those who want to buy it. It provides a straightforward mechanism for indexing large amounts of data, making it available, and advertising the availability.

The system is flexible enough to provide for a variety of billing methods. A small database maintainer might make the information available through a telephone connection. Using a 900 number, the billing would be taken care of by the phone company. A slightly more sophisticated site might have a password and credit card billing system. High volume servers might want to set up flat fee contracts with customers. Other methods will certainly emerge as use increases. The system was designed to be as adaptable as possible to future financial arrangements.

As the dissemination of information becomes easier, questions of ownership, copyright, and theft of data must be addressed. These issues confront the entire information processing field, and are particularly acute here. The WAIS system is designed to keep control of the data in the hands of the servers. A server can choose to whom and when the data should be given. Documents are distributed with an explicit copyright disposition in their internal format. This is not to say that theft cannot occur, but if a client starts to resell another's data, standard copyright laws can be invoked.

Directory of Servers

As the WAIS system develops, sources of information will proliferate, making it impossible for any user to keep track of all servers that may be available at any one time. To help solve this problem, Thinking Machines is maintaining a *Directory of Servers* in a widely accessible location. The Directory of Servers contains indexed textual descriptions of all known servers. It is queried just like any other source. Instead of text documents, however, it returns source structures, specially formatted files which can be plugged into a question and used for queries.

For example, suppose you needed information concerning the current gross national product of Mali, but had no idea where to find it. You might first ask the directory of servers for "information about the current economic condition of Mali." The directory would return several documents, among them might be a source for the *World Factbook,* an on-line almanac maintained by the CIA. You would then use this document as the source field of a question, and re-run the query. This time, the system would contact the almanac, ask for the information, and return a document with the data you need.

Additionally, the Directory of Servers provides a means for information providers to advertise the availability of their data. When a new source becomes available, the developers can submit a textual description, along with the necessary information for contacting the server. This information is added to the directory, and becomes available to the public.

A common protocol for information retrieval

One of the most far reaching aspects of this project is the development of an open protocol. The four companies have jointly specified a standard protocol for information retrieval. Creating a market where new servers can be readily established requires an open, publicly available protocol. Ideally this protocol would be internationally standardized, yet flexible enough to adapt to new ideas and technologies; functioning over any electronic network, from the highest speed optical connections to phone lines.

continued on next page

Wide Area Information Servers *(continued)*

The use of an open and versatile protocol fosters hardware independence. This not only provides for a much wider base of users, it allows the system to seamlessly evolve over time as hardware technology progresses. It provides incentive to produce the best components possible.

For example, the protocol provides for the transmission of audio and video as well as text, even though at present most workstations are unable to handle them. However, they are free to ignore pictures and sound returned in response to questions, and to display and retrieve only text. This inability, though, does not hinder higher-end platforms from exploiting their greater processing power and network bandwidth.

The WAIS protocol is an extension of the existing *Z39.50* standard from NISO [3]. It has been augmented where necessary to incorporate many of the needs of a full-text information retrieval system [4]. To allow future flexibility, the standard does not restrict the query language or the data format of the information to be retrieved. Nonetheless, a query convention has been established for the existing servers and clients. The resulting WAIS Protocol is general enough to be implemented on a variety of communications systems.

The success of a WAIS-like system depends on a critical mass of users and information services. In order to encourage development and use, Thinking Machines is not only publishing a specification for the protocol, but is also making the source code for a WAIS Protocol implementation freely available. While this software is available at no cost, it comes with no support. We hope that it will facilitate others in developing servers and clients.

Future

In developing the WAIS system, the participating companies have demonstrated that current hardware technology can be effectively used to provide sophisticated information retrieval services to novice end-users. How this might effect information providers is not yet completely understood. The users at Peat Marwick found the technology useful for day-to-day tasks such as researching potential new accounts and finding resources within their own organization. Since these tasks are not restricted to the accounting and management consulting industries, we are optimistic that this type of technology can be fruitful and productive in many corporate settings.

The future of this system, and others like it, depends upon finding appropriate niches in the electronic publishing domain. Potential uses include making current online services more easily accessible to end-users; or allowing large corporations to access their own internal word processor files more efficiently. It is also possible that near-term development will focus on a single professional field such as patent law or medical research.

Summary

A unique alliance of four companies with complementary interests in the field of information retrieval have jointly developed a prototype which gives versatile access to full-text documents. The system allows users to retrieve personal, corporate, and wide area information through one easy-to-use interface. The WAIS project has shown that current technologies can be used to make useful, profitable, and convenient wide area information systems. The success of the project has convinced us that a WAIS-like system can be a valuable tool for corporate information retrieval.

Acknowledgements

The design and development of the WAIS Project has been a collective effort, with contributions and ideas coming from many people. Among them:

Apple Computer: Charlie Bedard, David Casseras, Steve Cisler, Tom Erickson, Ruth Ridder, Eric Roth, John Thompson-Rohrlich, Kevin Tiene, Gitta Soloman, Oliver Steele, Janet Vratny-Watts. *Dow Jones News/Retrieval:* Clare Hart, Rod Wang, Roland Laird. *Thinking Machines:* Dan Aronson, Franklin Davis, Jonathan Goldman, Chris Madsen, Harry Morris, Patrick Bray, Danny Hillis, Gary Rancourt, Tracy Shen, Craig Stanfill, Steve Swartz, Ephraim Vishniac, David Waltz. *KPMG Peat Marwick:* Chris Arbogast, Mark Malone, Tom McDonough, Robin Palmer. *Scolex Information Systems:* Art Medlar. Thanks also to Advanced Software Concepts for *TCPack* software.

For more information

Brewster Kahle
Thinking Machines Corporation
1010 El Camino Real, Suite 310
Menlo Park, CA 94025
415-329-9300 x228
brewster@Think.com

Thinking Machines Corporation
245 First Street
Cambridge, MA 02142
617-234-1000

References

[1] Salton, Gerald; McGill, Michael. *Introduction to Modern Information Retrieval,* McGraw-Hill, 1983.

[2] *DowQuest* promotional literature available from Dow Jones & Co. Inc., 200 Liberty Street, New York, NY 10281.

[3] "Z39.50-1988: Information Retrieval Service Definition and Protocol Specification for Library Applications," National Information Standards Organization (Z39), P.O. Box 1056, Bethesda, MD 20817. (301) 975-2814. Available from Document Center, Belmont, CA. Telephone 415-591-7600.

[4] Franklin Davis et al., "WAIS Interface Protocol Prototype Functional Specification," Thinking Machines, Corp. Available from Franklin Davis (fad@think.com) or Brewster Kahle (brewster@think.com).

[5] Barron, Billy, "Another use of the Internet: Libraries Online Catalogs," *ConneXions,* Volume 5, No. 7, July 1991.

[6] Quarterman, John, "Networks: From Technology to Community," *ConneXions,* Volume 5, No. 7, July 1991.

[7] Schwartz, Michael F., "Resource Discovery and Related Research at the University of Colorado," *ConneXions,* Volume 5, No. 5, May 1991.

BREWSTER KAHLE is Project Leader for the Wide Area Information Servers project. With Thinking Machines since the company was founded in 1983, Brewster architected the CPU of the Connection Machine Model 2 and lead the design of all of the custom chips. For the last 2 years he has been working on making the supercomputers a smart information server in a joint project with Apple, Dow Jones and Peat Marwick. He can be reached as brewster@Think.com.

Scolex founder **ART MEDLAR** has been active in the Information Retrieval field for many years. He designed an early WAIS prototype on the Apple Macintosh, and supplied the telecommunications code for the current system. His company provides IR and telecommunications consultation services to clients throughout the San Francisco Bay area.

[Ed.: This article also appeared in the September 1991 issue of *ONLINE Magazine.*]

Multimedia Mail From the Bottom Up
or
Teaching Dumb Mailers to Sing

by Nathaniel S. Borenstein, Bellcore

Abstract

Multimedia mail systems have exhibited great potential, but the widespread use of multimedia mail has so far been inhibited by the lack of interchange standards and the heterogeneity of mail-reading software. This article describes a new approach that seeks to break the existing log-jam and make multimedia mail a practical reality. The article begins with a brief summary of the state of the art in multimedia mail systems. It then outlines the new, "bottom-up" approach, and describes the configuration mechanism that is central to its operation. The article ends by outlining a vision of a new and better "lowest common denominator" for electronic mail.

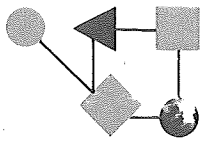
The promise of multimedia mail

Electronic mail (e-mail) is a widely-used and much-appreciated technology. Ever since the inception of electronic mail, there has been much discussion of its even greater potential. For most people, e-mail today is a text-only medium, in which unformatted textual messages can be sent rapidly to even the most distant of correspondents. In principle, the limitation to plain text is artificial. E-mail is fundamentally capable of carrying richly formatted text, images, audio, video, and indeed anything that can be encoded in a digital form. In practice, however, the vast majority of the world's e-mail users are still restricted to plain text, due to a lack of interchange standards and a profusion of heterogeneous software for reading mail. The relatively few users of advanced multimedia mail systems such as *The Andrew Message System* [1] and *Diamond* [5] can only interchange multimedia mail with other users of the same software. An *Andrew* user and a *Diamond* user cannot, for example, send mail with pictures to each other. The result is that no multimedia mail technology has reached "critical mass" and made anything beyond plain text a part of the standard e-mail infrastructure for the masses.

The approach taken by most multimedia mail system to date can be characterized as a "top-down" approach. The developers of such systems said to their potential users, like Moses coming down from Mount Sinai, "Behold! I give you multimedia mail. All you need to do, in order to reap its blessings, is to change your mail reading program, your mail sending program, your text editor, your drawing editor, and generally everything about the way you work on a computer. Oh, and all your correspondents must do the same." When viewed in this way, it is perhaps not surprising that the world has not rushed headlong to embrace any of these systems.

The situation is best illustrated by considering the two different types of sites where *Andrew* is in use. At some sites, including the Carnegie Mellon University campus, where *Andrew* was developed, its use is nearly ubiquitous. (This was typically accomplished by administrative fiat.) Given this fact, the sender of a message can rely on the ability of the recipients to see a multimedia message in all its splendor. In such environments, a substantial portion of all mail messages contain at least multi-font text, and mail containing images, hypertext links, or other multimedia objects is not uncommon. At the other extreme, however, are sites where only a few individuals have elected to use *Andrew*.

CONNEXIONS



The Interoperability Report

November 1991 Volume 5, No. 11

ConneXions —
The Interoperability Report tracks current and emerging standards and technologies within the computer and communications industry.

In this issue:

ConneXions is published monthly by Interop, Inc., 480 San Antonio Road, Suite 100, Mountain View, CA 94040, USA. 415-941-3399. Fax: 415-949-1779. Toll-free: 1-800-INTEROP.

ISSN 0894–5926

From the Editor

After two special issues focusing on INTEROP 91 Fall, it is time to return to "normal" and catch up with some of the topics *not* directly related to the show which got pushed aside in the last couple of months. Of course, we will return to INTEROP 91 Fall in a future issue (most likely December 1991) with reports and pictures, so stay tuned.

Our July issue, subtitled "The Changing Face of the Internet," looked at a number of new and interesting applications of Internet technology. This month, we continue this thread with a look at WAIS, Multimedia Mail, and Resource Discovery.

The *Wide Area Information Server* (WAIS, pronounced "ways") project is an experimental venture seeking to determine whether current technologies can be used to make profitable end-user full-text information systems. Our first article, written by Brewster Kahle and Art Medlar, discusses the design and implementation of the prototype WAIS system.

Multimedia mail systems have actually been in use on the Internet and elsewhere for many years. However, no multimedia mail technology has reached critical mass, due in part to the variety of interchange standards and systems in use. Nathaniel Borenstein of Bellcore gives a brief summary of the state of the art in multimedia mail systems. The article describes a new "bottom-up" approach to multimedia mail, and outlines a vision of a new and better "lowest common denominator" for electronic mail.

In a recent study, researchers at the University of Colorado, involved with the Resource Discovery project, attempted to measure the nature of connectivity to the Internet by sending certain simple "probes" to a statistical sample of host. The reaction to this experiment is the subject of an article by Carl Malamud on page 18. It should be noted that the IAB recently issued a statement—in the form of RFC 1262—on the subject of Internet Measurement. The summary is included below:

"Measurement of the Internet is critical for future development, evolution and deployment planning. Internet-wide activities have the potential to interfere with normal operation and must be planned with care and made widely known beforehand. This document offers guidance to researchers planning Internet measurements. This RFC represents IAB guidance for researchers considering measurement experiments on the Internet. This RFC does not represent a standard for the Internet but the Internet Activities Board strongly urges that Internet users follow the guidelines out of courtesy and professional consideration for the Internet community."

3/81

An Information System for Corporate Users: Wide Area Information Servers

by
Brewster Kahle, Thinking Machines Corporation
and
Art Medlar, Scolex Information Systems

Background

To explore text-based information systems for corporate executives, four companies have jointly developed a prototype which gives flexible access to full-text documents. The four participating companies are Dow Jones & Co., with its premier business information sources; Thinking Machines Corporation, with its high-end information retrieval engines; Apple Computer, with its user interface expertise; and KPMG Peat Marwick, with its information-hungry user base.

One of the primary objectives of the project is to allow a user to retrieve personal, corporate, and wide area information through one easy-to-use interface. For example, instead of using Lotus *Magellean*™ for personal information, *Verity Topic*™ for corporate data, and *Dialog*™ for published text, one application can access all three categories of information. The user isn't required to become familiar with several entirely different systems. In addition, since the interface consolidates data from many different sources, they can be manipulated effortlessly, virtually without regard to their origins.

The *Wide Area Information Server* (WAIS, pronounced "ways") project is an experimental venture seeking to determine whether current technologies can be used to make profitable end-user full-text information systems. Fifteen users have been actively using the system for over three months. They have integrated it into their workday routine in much the same way as they have previously integrated spreadsheets and word processors. This preliminary success has convinced us that a WAIS-like system can be a valuable tool for corporate information retrieval. This article discusses the design and implementation of the prototype system.

Introduction

Electronic publishing is the distribution of textual information over electronic networks. It has been emerging as a viable alternative to traditional print publishing as the necessary underlying technologies develop. Among the more essential of these are:

- High Resolution Display Screens
- Reliable, High-Speed Data Communications
- Desktop Publishing Systems
- Inexpensive Data Storage Media

While these technologies have been developed for uses other than electronic publishing, they are the necessary precursors for full-text retrieval systems.

From the user's point of view, there are several problems to be overcome. First, there must be some way of finding and selecting databases from a potentially unlimited pool. Second, although these databases may be organized in different ways, the user should not need to become familiar with the internal configuration of each one. Finally, there must be some practical way of organizing responses on the users machine in order to maintain control over what may become a vast accumulation of data.